SSCI 583 Spatial Analysis
Project 1 Report

Group member: Yongchun Chen, Junhong Duan, Shiqi Li, Ruochen Liang, Lingduo Luo

1 Introduction

The goal of this project is to select suitable locations for building new healthcare facilities, such as hospitals, clinics, or medical offices, in the San Francisco area. Using Geographic Information Systems (GIS) and multiple criteria decision analysis (MCDA) methods with ArcGIS Pro, we choose seven criteria, such as distance to main roads, population density, and slope to conduct both weighted overlay and fuzzy overlay.

Data for the project all come from DataSF, including boundary, DEM raster, land use parcels, and demographic information from the census. Weighted overlay and fuzzy overlay are two methods used in GIS spatial analysis to combine multiple layers of spatial data to create a new layer that represents a specific analysis or decision-making scenario. In this project, we use the two methods to weigh the relative importance of different criteria to the siting of a healthcare facility to give scores of different areas and identify suitable locations.

Overall, this project typically involves researching and analyzing areas of need, such as neighborhoods or communities with limited access to health care or high rates of certain medical conditions. By selecting suitable locations, the health care facilities can ensure that the people who need it can access the care.

2 Study area

The study area of this project is mainland San Francisco County, which is located in the northern part of California, United States. It is the most densely populated county in California and the second-most densely populated county in the United States with a population of over 883,000 people.[1] The county is home to a number of major healthcare providers, including the University of California, San Francisco Medical Center, California Pacific Medical Center, and the San Francisco General Hospital.

San Francisco has a high population density, which means that there is a large potential patient base for new healthcare facilities. Additionally, due to its high cost of living and high real-estate cost, it could be challenging to select suitable sites to build healthcare facilities. Taking into account the criteria of a location, an appropriate site selection for building a new healthcare facility could be beneficial to a community in San Francisco.

Fig. 1 Study Area

3 Criteria

The goal of this project is to choose the suitability of location for healthcare facilities, which can vary depending on the specific needs of the community and the facilities itself. Referring to the findings of Dell'Ovo's research[2], our group came to an agreement of an evaluation system that divided the criteria into four categories: Functional quality, Location quality, Environmental quality, and Economical aspects.

Besides the four categories, a prerequisite criteria was applied to the research, which is the proper land use parcel. According to the land use data documentation and the report from the American Planning Association (APA)[3], the facilities should be located in a parcel where the necessary zoning regulations and permits can be obtained. We eliminated the industrial, open space, and right of way land use parcels and limited the parcel's area to 1,500 square feet to 30,000 square feet to exclude parcels of inappropriate size.

Functional quality refers to the ability of a healthcare facility to provide necessary services and meet the needs of patients. In this project, we choose the population density and population income as the representation of functional quality.

Location quality refers to the accessibility and proximity of the healthcare facility to the population it serves. We choose the distance to main roads, distance to current healthcare facilities and slope for this type of criteria.

Environmental quality refers to the physical environment of the healthcare facility. In this project, we chose noise pollution as a representation.

Economical aspects refers to the financial viability of the healthcare facility, including the value of the area and the land ownership. Unfortunately, the economical aspects are usually inaccessible and need detailed evaluation and negotiation. So in this project, we discard this part of the criteria.

The criteria in the study area are as Fig. 2 - Fig. 7.

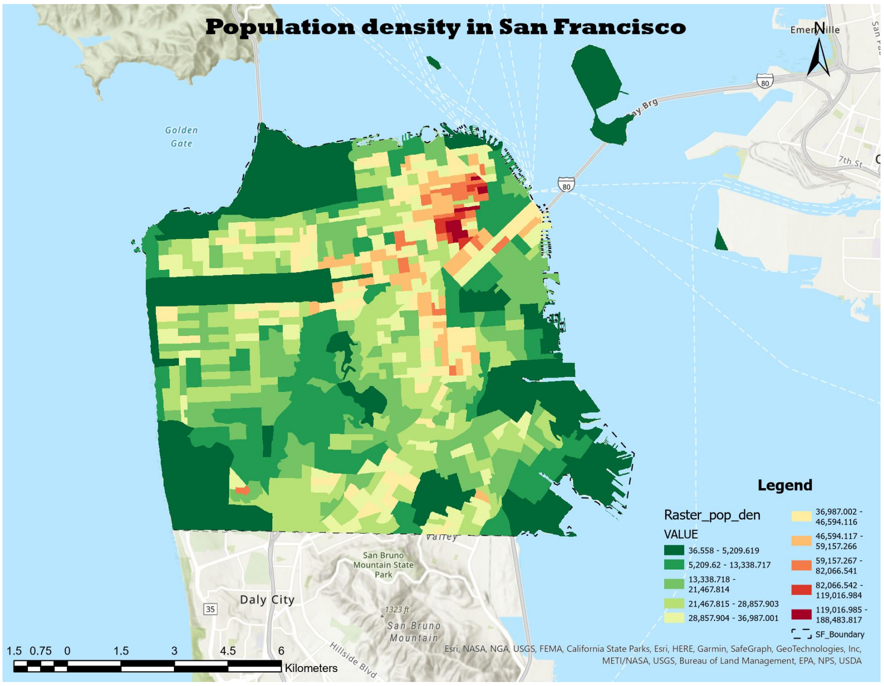

Fig. 2 Population Density Raster

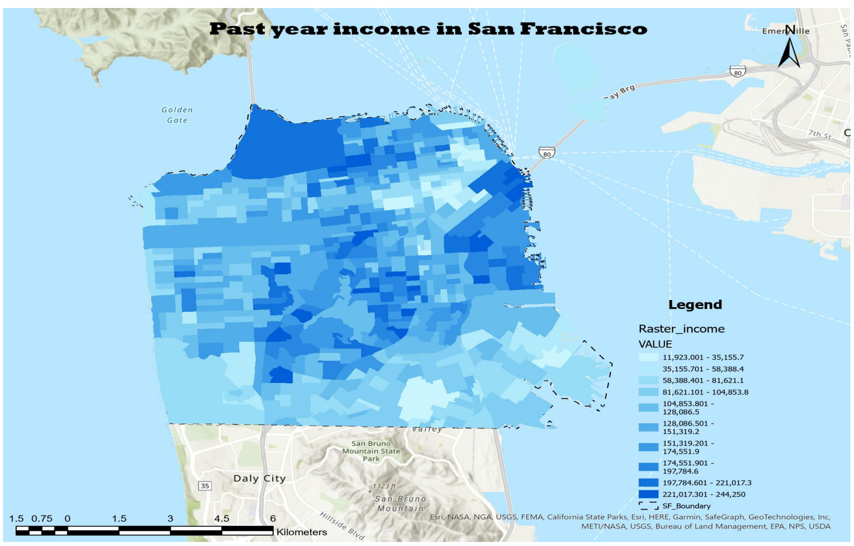

Fig. 3 Population Income Raster

Fig. 4 Slope Raster

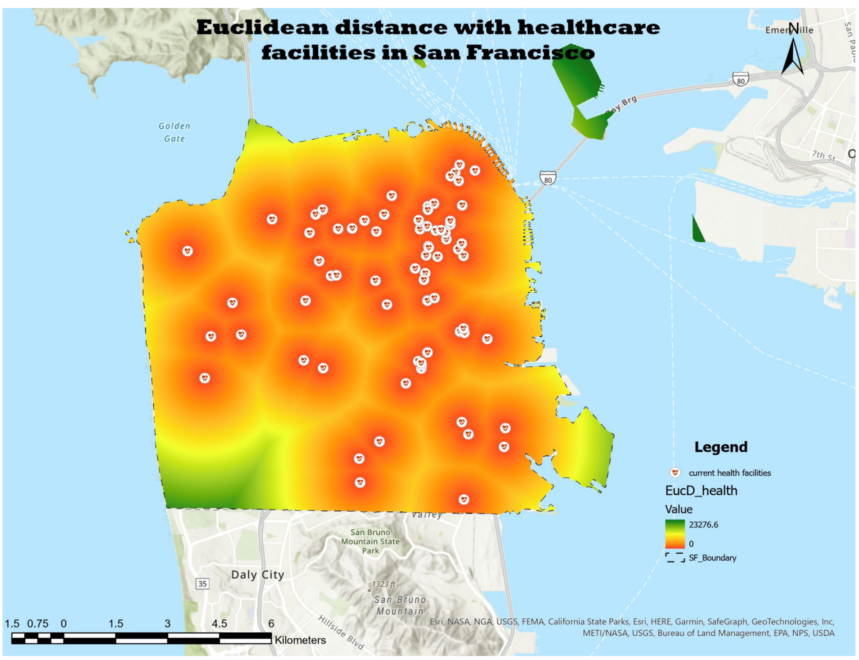

Fig. 5 Distance From Current Healthcare Facilities Raster

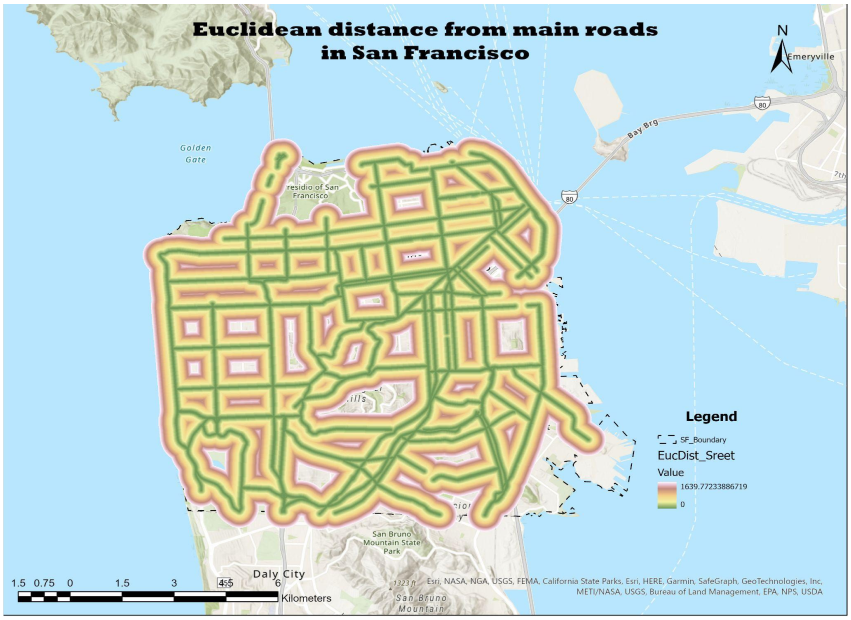

Fig. 6 Distance From Main Roads Raster

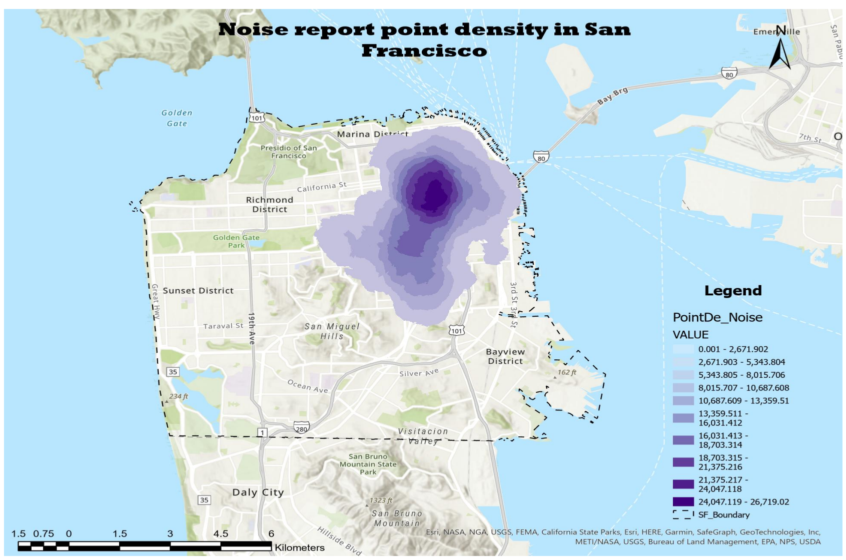

Fig. 7 Noise Pollution Raster

The reasons for choosing the criteria are explained as table 1.

Table 1 Criteria choices

Criteria	Categories	Reason
Population density	Functional Quality	In areas with high population density, there is typically a greater need for healthcare services and facilities, as a higher number of people means a higher demand for healthcare services.
Population	Functional	In areas with higher income populations, healthcare facilities may be

income	Quality	more likely to be small private clinics that offer more specialized and higher-cost services, such as cancer centers or specialty clinics. Under the same population density, the high-income population area has greater demand for those healthcare facilities.
Distance to main roads	Location Quality	Being close to main roads can make the facility more accessible to patients, especially for those who are traveling by car. This can also be beneficial for emergency medical services, as they can reach the facility more quickly.
Distance to current healthcare facilities	Location Quality	New healthcare facilities located near existing ones may face competition from the existing facilities, which could lead to decreased revenue and financial instability for the new facility.
Slope	Zoning regulations	Building on a slope can pose safety risks, particularly if the slope is steep. Construction crews must take special precautions to prevent landslides or other hazards that may occur during the construction phase or even after the facility is completed. Also, it adds additional costs to the construction of a healthcare facility. Additional infrastructure, such as retaining walls or drainage systems, may be required to ensure the safety and stability of the facility.
Noise pollution	Other	Noise level can be an important factor to consider when choosing a location for healthcare facilities, as it can affect the well-being of patients and staff. High levels of noise can lead to increased stress, sleep disturbances, and other negative health effects, which can impact the recovery of patients and the overall functioning of the healthcare facility.

4 Data table

Table 2 Data Table

Dataset	**Criterion used for**	**Source of data**
Bay Area Counties[4]	Boundary	DataSF
Census 2020: Block Groups for San Francisco[5]	Population density, population income	DataSF
Land Use of San Francisco[6]	Pre-screening parcels	DataSF
Health Care Facilities of San Francisco[7]	Distance to existing healthcare facilities	DataSF
Arterial Streets of San Francisco[8]	Distance to main streets	DataSF

San Francisco Bay-Delta DEM[9]	Slope	Data.gov
311 Cases[10]	Noise Pollution	DataSF

5 Methods

5.1 Data Preparation

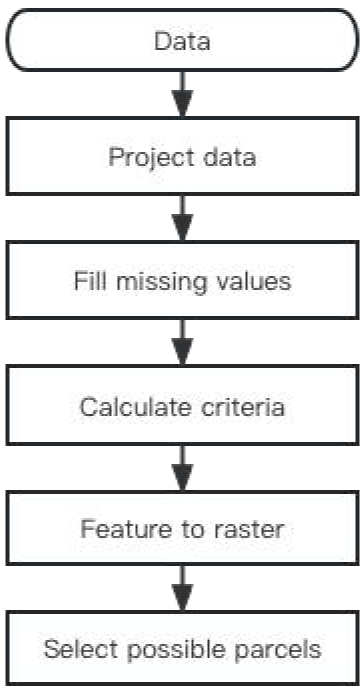

Fig. 8 Data Preparation Workflow

1. Project data into unified spatial coordinates

To make sure that all the data used in the analysis is in the same coordinate system, we use the “Project” tool in ArcGIS Pro to project all the data into NAD 1983 (2011) StatePlane California III FIPS 0403 (US Feet), same as the boundary data. This is important for overlay analysis and other spatial operations to work correctly.

2. Fill missing values

This step involves identifying and filling in any missing values in the data. It was done with the “Fill Missing Values” tool in ArcGIS Pro. Specifically, those data had missing values and were filled as follows. The missing value of the population density and population income was filled with the average values from 5 nearest neighbors.

3. Calculate criteria

Some criteria should be calculated with tools in ArcGIS Pro for the analysis.

In this project, the population density was calculated by dividing the total population by the area of the block group using the “Calculate Geometry” and “Calculate Field” tools.

The slope was calculated using the “Slope” tool with DEM data.

Distance from the main road and distance from current healthcare facilities were calculated by the “Euclidean Distance” tool. When calculating the distance from the main road, the maximum distance was set to 500m(1640 feet), because the furthest distance between current healthcare facilities in SF and the main road is about 500m, which should be the threshold in our analysis. When calculating the distance from current healthcare facilities, the maximum distance was set as 25 kilometers, which is the specialized hospital like a children's hospital or a trauma center average catchment area.

Noise pollution was first used “XY Table to Point” to convert from a CSV file to a feature layer. Then we used the “Point Density” tool to calculate the density of the noise-reported points.

4. Feature to raster

This step contains converting the vector data we collected to raster data. This was done using the “Feature to Raster” tool in ArcGIS Pro. For the scale of the output raster data, we decided on 10m x 10m because it’s the same as the smallest grid size of our data, and it is near the smallest medical parcels in the land use data.

5. Select possible parcels

According to chapter 3, we select land use parcels between 1,500 and 30,000 areas that are not for Industrial use, open space, and right of way. We use the “Select Layer By Attributes” tool to pick the possible parcels as figure 8. The mixed parcels are concentrated in downtown San Francisco. The rest of the possible parcels are mainly residential land use parcels. The selected possible parcels are as Fig. 9

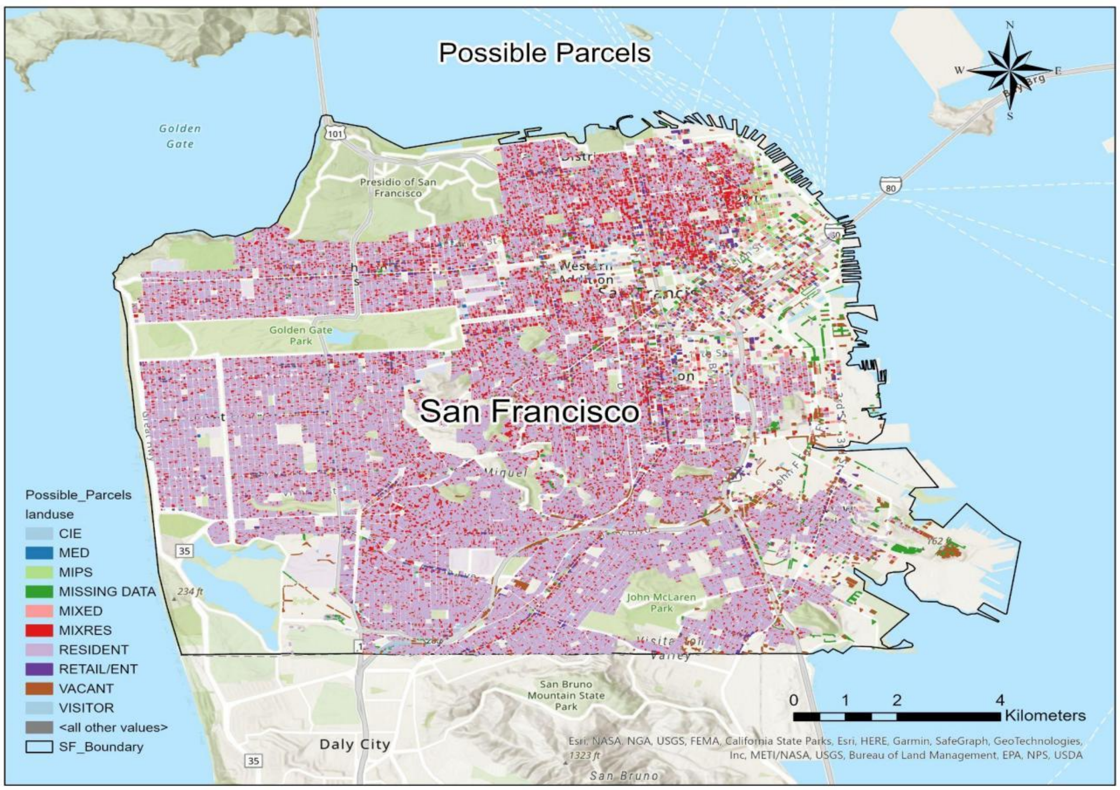

Fig. 9 Possible Parcels After The Pre-requisite Filtering

5.2 Weighted Overlay

In GIS, the weighted overlay is a technique for combining multiple raster layers into a single composite map, with each layer representing a distinct criterion of a decision problem. By first reclassifying a number of input rasters into a single measurement and then multiplying each raster with an assignment percentage influence, the weighted values of the result cells are added together.[11]

Fig. 10 Workflow of The Weighted Overlay

1. Data Reclassification

The weighted overlay needs the raster data to be reclassified in advance. We choose the scale for reclassification as one to five and the cutoff method as natural breaks for these are the commonly selected scale and methods.

We gave higher scores to areas with higher population density, higher median income, closer to main roads, further from current healthcare facilities, smaller slopes, and less noise pollution. Lower scores were given to the opposite grids.

2. Weighted Overlay

Combining the literature review and our discussion, we gave weights for each criterion as table 3. After choosing the weighting scheme, we use the “Weighted Overlay” tool to conduct the overlay.

Table 3 Weighting Scheme for Weighted Overlay Analysis

Criteria	**Weight**
Distance to current healthcare facilities	30%
Population density	25%
Population income	20%
Distance to main roads	15%
Noise pollution	5%
Slope	5%

5.3 Fuzzy Overlay

The Fuzzy Overlay method allows the analysis of the possibility of a phenomenon belonging to multiple sets in a multicriteria overlay analysis. Not only does Fuzzy Overlay determine what sets the phenomenon is possibly a member of, but it also analyzes the relationships between the members of the multiple sets. Our workflow is shown below (Figure 10).

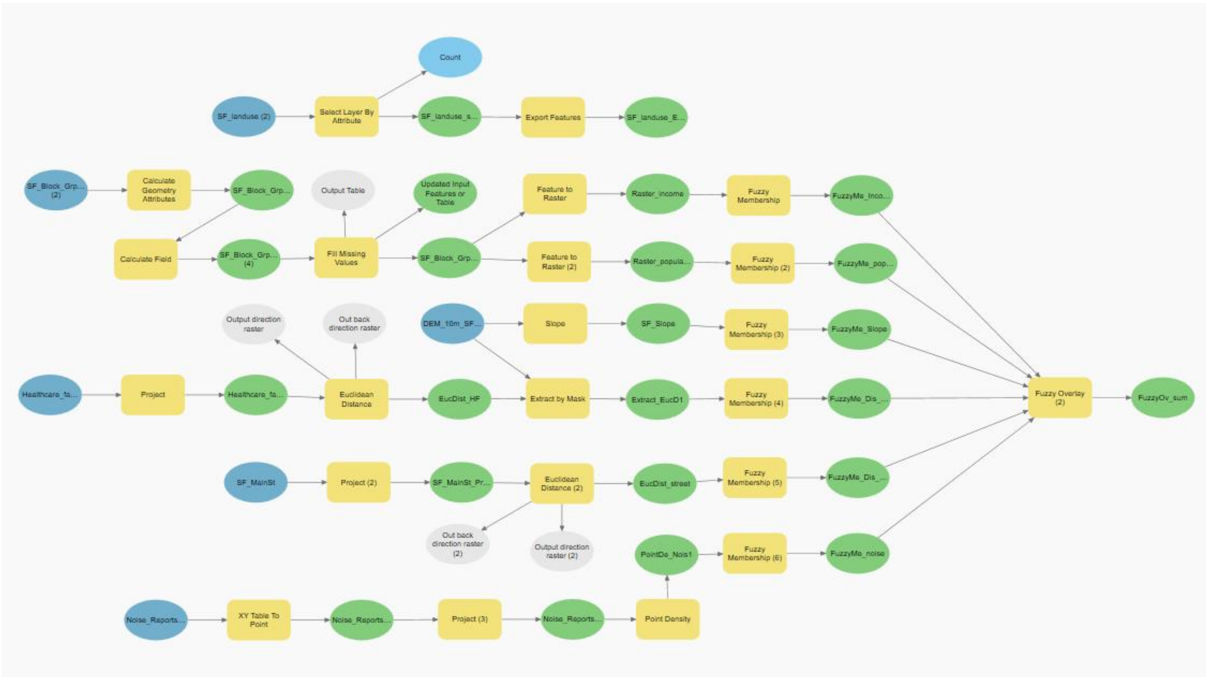

Fig. 11 Workflow for Fuzzy Overlay Analysis

1. Create fuzzy membership sets

Fuzzy Membership Sets are required before choosing a Fuzzy Overlay analysis so that the raster grid value will be set between 0-1. We use the 6 layers after rasterization to run Fuzzy Membership one by one. In the Fuzzy membership tool, two main parameters are introduced: membership type and spread. In this project, we set parameters as follows (table 4). Figure 11 shows the result of the Near function for Distance to other healthcare facilities:

Table 4 Parameters for fuzzy membership tool

Raster layer	Membership type	spread	reason
Slope	small	Default 5	High raster value represents high slope, which should have a small value.
Population density	large	Default 5	High raster value represents high density, which should have a large value.
Noise points	small	Default 5	High raster value represents high density of noise, which should have a small value.

Income by block_Grps	large	Default 5	High raster value represents high income, which should have a large value.
Distance to Healthcare facilities	Near	Default midpoint	Since we would like to choose locations neither too far nor too close to other healthcare facilities, Near type allows us to set higher values for midpoints. In Figure 11, yellow color has higher value and dark green has lower value.
Distance to main street	small	Default 5	High raster value represents long distance, which should have a small value.

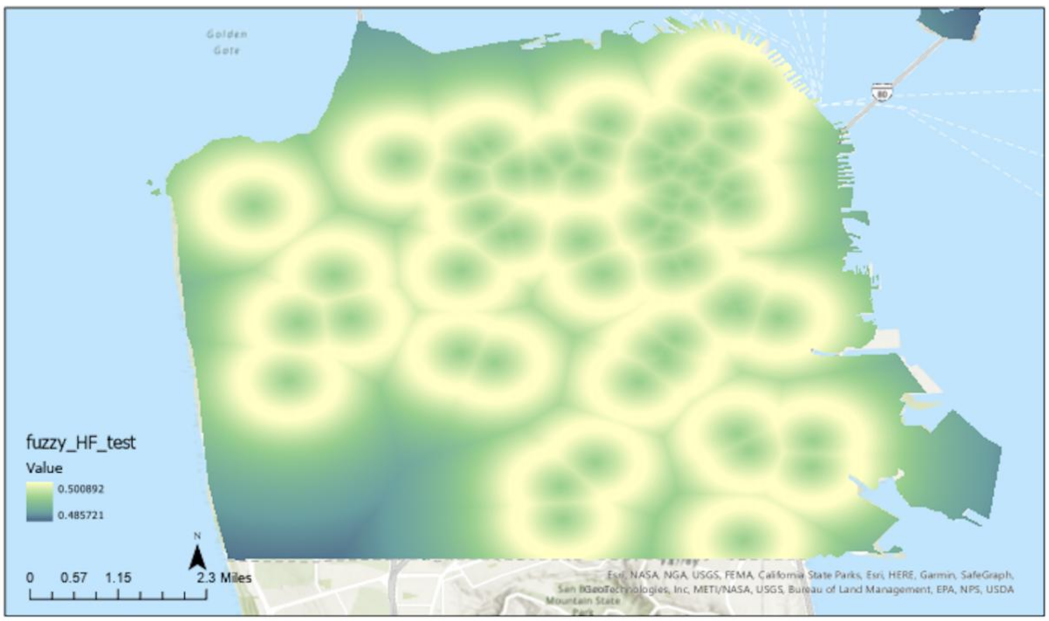

Fig. 12 Fuzzy distance to other healthcare facilities using Near function in Fuzzy Membership tool

2. 2.Choose and apply Fuzzy overlay method

A Fuzzy Overlay in ArcGIS Pro results in a raster map where each cell is assigned a fuzzy membership value based on the input maps. This value represents the degree of belonging of each cell to each class. In the Fuzzy overlay tool, there are many different operators of Fuzzy Overlay that could be chosen, such as fuzzy And, Or, and Product overlay type. Based on the criteria we selected, all the criteria have a combined influence on suitability analysis. The fuzzy “AND" overlay type will return the minimum value of the sets the cell location belongs to. This technique helps to identify the least common denominator for the membership of all the input criteria. Therefore, we finally decided to use Overlay "And".

6 Results

6.1 Results for Weighted Overlay

Combine the layers by multiplying the value of each cell in each layer by the corresponding weight and summing the results for all layers. The output map has the mask of “land selection”, which are the most suitable areas for specific land use, the dark green places are the potential locations for building the healthcare facilities.

According to the results, the most suitable results are located in the Marina District, the west part of Richmond District, the South part of Japantown, the middle-west part of Mission District, the west part of Sunset District, and the east part of Parkmerced. All the suitable parcels are near the main road, which is the effect of the distance to the main road criterion. The most suitable parcels are not around downtown San Francisco because we use the distance to current healthcare facilities as one of the criteria, but there are still some moderate parcels in the downtown area, claiming the downtown area as a suitable location for the current healthcare facilities.

Therefore, we recommend focusing on the Marina District, Richmond District, and Mission district for building new healthcare facilities to fill the possible vacancies in healthcare of these communities.

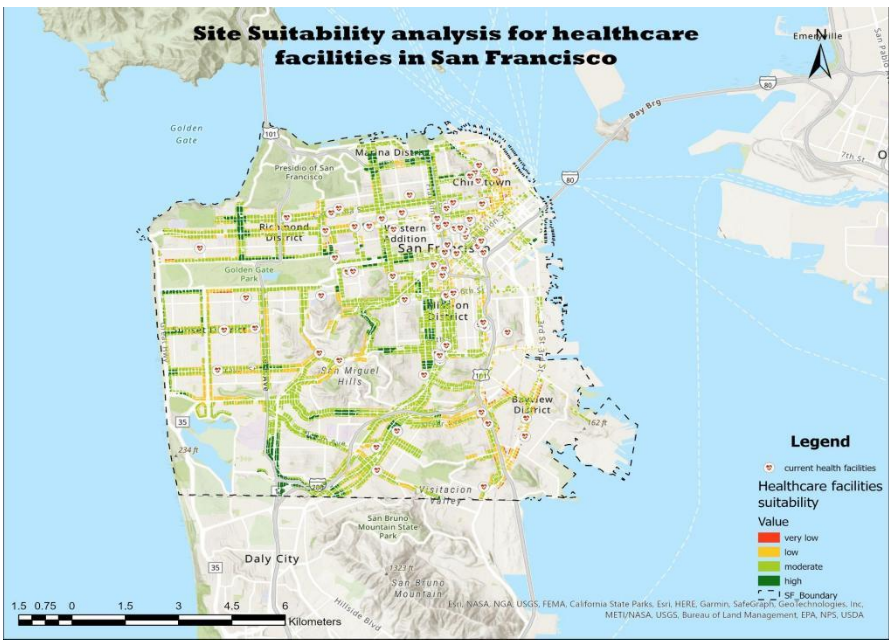

Fig. 13 Healthcare facilities suitability results using weighted overlay

6.2 Results for Fuzzy Overlay

In Fuzzy overlay analysis, we firstly get fuzzy membership sets and overlay them by Fuzzy Overlay. The final overlay map displays a continuous range of values from 0 to 1, indicating the level of membership of each cell to each class. This result map can be used to visualize and analyze the spatial relationship between raster layers and to make decisions based on the degree of membership of each cell to each class. In this project, Fuzzy And overlay type allows us to select locations which have high values under all the criteria.

The output map is shown in Figure 12. Dark green color represents more suitable locations, and yellow color represents less suitable locations. According to the result, we could learn that northwestern and southeastern parts (westRichmond district, south Sunset district, west Marina district, south Bayview district and west to the John McLaren Park) of San Francisco are more suitable for a new healthcare facility based on our criteria.

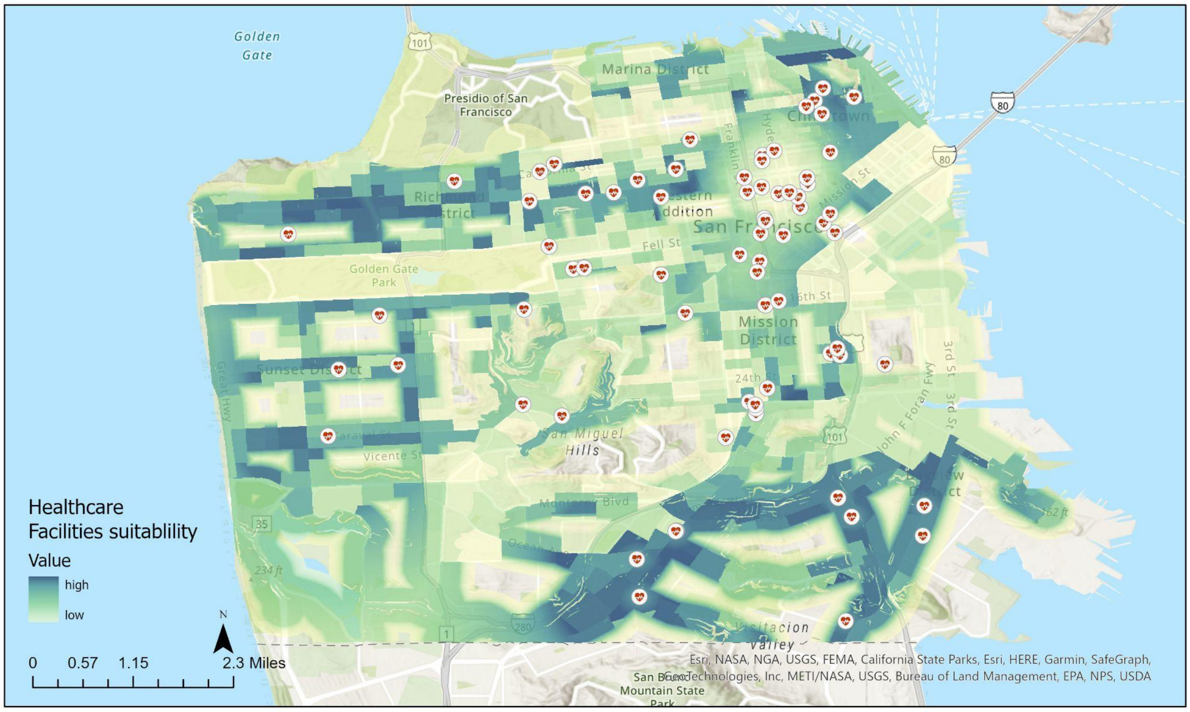

Fig. 14 Healthcare facilities suitability results using fuzzy overlay

6.3 Comparison of Two Analysis Results

It can be seen that the results obtained by fuzzy overlay are highly centralized. Focused on four main areas of San Francisco, the results of fuzzy overlay are mainly distributed at the Golden Gate Park area, the western part of the city near the sunset district area, the northeastern Chinatown area, and the southeastern Bernal heights area.

On the other hand, the results obtained from the weighted overlay are more dispersed. Mission district is more distributed, but also more dispersed, and the southwest area of Sanban City is also more suitable.

A more detailed analysis are focusing on the four categories of the criteria. For the functional quality, an overlay analysis with the population income distribution shows that the fuzzy overlay is more thoughtful than the weighted overlay in terms of the distribution of healthcare facilities for lower-income groups. This is evidenced by the fact that the sites recommended by fuzzy overlay are more likely to cover neighborhoods with a population income of less than $81,621 per year. In contrast, the weighted overlay sites are primarily located in neighborhoods with mid-class per capita incomes.

The two overlay methods have similar results in terms of population density distribution and site selection. Both of them take the population distribution into account to some extent, and both of them locate the sites near the neighborhoods where the population distribution is in the middle range.

Based on the existing hospital addresses, the results obtained by fuzzy overlay overlap to some extent with the existing hospital addresses. For example, the proposed distribution of major hospitals in the southeast region of the city overlaps with the existing medical buildings. The weighted overlay avoids the existing medical buildings to a certain extent. It is speculated here that the possible reason for this is that the influence of topography here is more reflected in the results of fuzzy overlay due to the weighting in the data criteria setting. Therefore, after more consideration of slope, sites on hillsides are considered more unsuitable areas in the fuzzy overlay.

Similarly, for the environmental quality, in the Chinatown area of northeastern San Francisco, where high levels of noise pollution and the presence of several existing medical facilities are present, both fuzzy overlay's site selection results and weighted overlay's site selection results avoid the area to some extent.

It is obvious from the resulting graph that the weighted overlay has a higher requirement for sitting near the main road. On the contrary, the fuzzy overlay has a certain choice for both the main road and the buffer area around the main road, and its preference for the main road is not as obvious as that of the weighted overlay. However, in general, the main recommended site selection points of both methods are concentrated in the areas on both sides of the main roads.

7 Discussion

Considering the criteria we selected, the actual situation, some subjective judgments, and the comparative analysis of weighted overlay and fuzzy overlay, our site selection for healthcare facilities in San Francisco is relatively complete. However, there are still some areas to be improved in our work.

One critical aspect concerns the selection of criteria. A sitting for a healthcare facility should take a large number of other factors into consideration. For example, the local green land area, local air quality and surrounding building density. However, due to the difficulty in obtaining data and other reasons, in the end we only selected six influencing factors. Apart from the quantity of criteria, when we analyzed and calculated the distance to current healthcare facilities and distance to main roads, we did not consider the issue of reachability, but directly used the Euclidean distance, which is not consistent with the actual situation and may have some impact on the accuracy of the results.

Besides, the Economic Dimension has not been considered, but that should have been taken into consideration in a site selection process. In fact, once the total area of the healthcare facility is determined, the site selection should pass through the analysis of Land value, Land Ownership and Infrastructural quality with the support of the MCDA.[2]

Additionally, this decision making process may involve various stakeholders such as healthcare providers, government officials, and community representatives. They may have different priorities and objectives, which need to be taken into account when choosing the best location. But in this project, we are only concerned about the geographical impact and population demographics data through ArcGIS Pro spatial analysis tools to find potential spaces.

In conclusion, site selection is a complicated problem, in which many fields and factors other than spatial and environmental factors should be considered. In future research, we should follow the official standards and consider the selection of criteria more carefully to improve the accuracy of site selection results.

8 Reference

[1] “San Francisco.” Wikipedia. Wikimedia Foundation, January 28, 2023. https://en.wikipedia.org/wiki/San_Francisco.
[2] Dell’Ovo, Marta, Stefano Capolongo, and Alessandra Oppio. “Combining Spatial Analysis with MCDA for the Siting of Healthcare Facilities.” Land Use Policy 76 (2018): 634–44. https://doi.org/10.1016/j.landusepol.2018.02.044.
[3] “Zone Locations for Hospitals and Other Medical Facilities.” American Planning Association. Accessed January 28, 2023. https://www.planning.org/pas/reports/report50.htm.
[4] City and County of San Francisco. “Bay Area Counties: DataSF: City and County of San Francisco.” San Francisco Data, July 13, 2016. https://data.sfgov.org/Geographic-Locations-and-Boundaries/Bay-Area-Counties/s9wg-vcph.
[5] City and County of San Francisco. “Census 2020: Block Groups for San Francisco: DataSF: City and County of San Francisco.” San Francisco Data, July 22, 2022. https://data.sfgov.org/Geographic-Locations-and-Boundaries/Census-2020-Block-Groups-for-San-Francisco/24e8-pd2q/data.
[6] City and County of San Francisco. “Land Use: DataSF: City and County of San Francisco.” San Francisco Data, April 25, 2016. https://data.sfgov.org/Housing-and-Buildings/Land-Use/us3s-fp9q.
[7] City and County of San Francisco. “Health Care Facilities: DataSF: City and County of San Francisco.” San Francisco Data, February 12, 2016. https://data.sfgov.org/Health-and-Social-Services/Health-Care-Facilities/jhsu-2pka/data.
[8] City and County of San Francisco. “Arterial Streets of San Francisco: DataSF: City and County of San Francisco.” San Francisco Data, July 28, 2016. https://data.sfgov.org/Geographic-Locations-and-Boundaries/Arterial-Streets-of-San-Francisco/gnsq-9x5h.
[9] Department of the Interior. “San Francisco Bay-Delta Bathymetric/Topographic Digital Elevation Model(Dem).” Catalog. Publisher U.S. Geological Survey, October 13, 2022. https://catalog.data.gov/dataset/san-francisco-bay-delta-bathymetric-topographic-digital-elevation-modeldem.
[10] City and county of San Francisco. “311 Cases: DataSF: City and County of San Francisco.” San Francisco Data, October 8, 2011. https://data.sfgov.org/City-Infrastructure/311-Cases/vw6y-z8j6/data.
[11] Esri ArcGIS Pro team. “Weighted Overlay (Spatial Analyst).” Weighted Overlay (Spatial Analyst)-ArcGIS Pro | Documentation. Accessed January 29, 2023. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/weighted-overlay.htm.

9 Individual Contribution to the overall project

Name	Contribution
Yongchun Chen	Data Searching; Criteria Selection; Literature Review; Operating of Fuzzy Overlay; Report Writing of Discussion
Junhong Duan	Literature Review; Criteria Selection; Operating of Fuzzy Overlay; Report Writing of Fuzzy Overlay Method and Results
Shiqi Li	Task Distribution; Literature Review; Workflow Presentation; Report Writing of Study Area, Data Table, and Weighted Overlay Resul; Report Integration and Modification
Ruochen Liang	Literature Review, Operating of Data Preprocessing and Weighted Overlay, Report Writing of Introduction, Criteria, Weighted Overlay Method
Lingduo Luo	Data Searching; Criteria Selection; Operating of Fuzzy Overlay; Report Writing of Fuzzy Overlay Method and Comparison of two methods